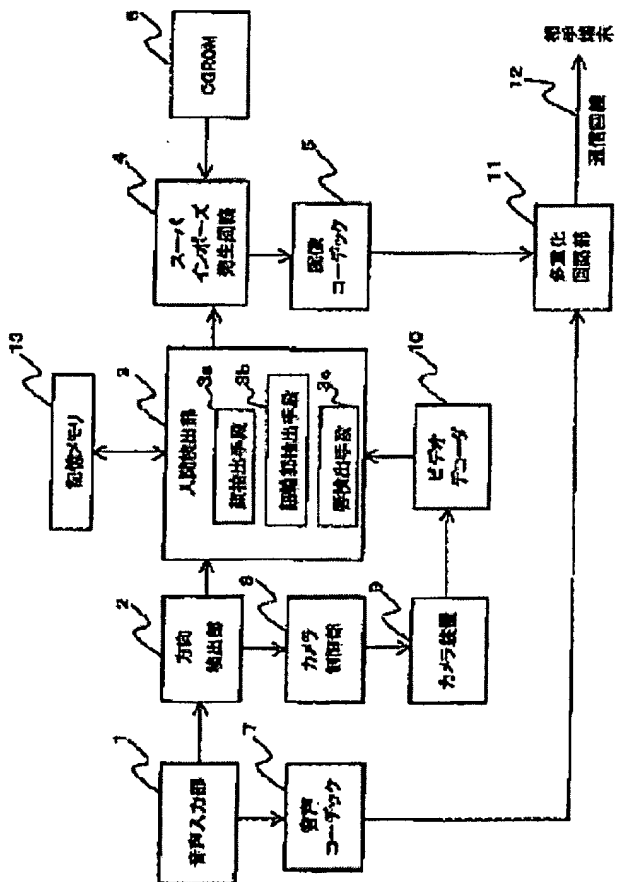


**SPEAKER IDENTIFYING DEVICE AND VIDEO CONFERENCE SYSTEM PROVIDED WITH SPEAKER IDENTIFYING DEVICE**

<b>Patent number:</b>	JP2003189273
<b>Publication date:</b>	2003-07-04
<b>Inventor:</b>	IMAI TAKASHI; IWASAKI KAZUYA
<b>Applicant:</b>	SHARP KK
<b>Classification:</b>	
- international:	H04N7/15; G06T1/00; G06T7/00; G06T7/20
- european:	
<b>Application number:</b>	JP20010387569 20011220
<b>Priority number(s):</b>	JP20010387569 20011220

# Abstract of JP2003189273

<P>PROBLEM TO BE SOLVED: To provide a video conference system capable of always and accurately displaying a speaker on a monitor screen. <P>SOLUTION: If a human detecting part 3 for detecting a person detects a person in an input image in a voice source direction obtained by a direction detecting part 2 for detecting the voice source direction of an input voice, a CGROM 6 and a superimposition generation circuit 4 display a prescribed marker indicating a speaker around the image of the detected person on the monitor screen. The human detecting part 3 is provided with a face extracting means 3a for extracting a face color of a person to detect the person, a facial contour detecting means 3b for detecting the facial contour, eyes, nose and lips of the person, a lips detecting means 3c for detecting the speaker by detecting lips' movement. The direction detecting part 2 can be provided with a first storage memory for storing voice source characteristics inputted in advance as a video conference participant, and a voice source characteristics comparing means for comparing the characteristics of a newly inputted voice source with those of the voice source stored in the first storage memory. <P>COPYRIGHT: (C)2003,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-189273

(P2003-189273A)

(43) 公開日 平成15年7月4日 (2003.7.4)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テグト <sup>*</sup> (参考)
H 0 4 N 7/15	6 3 0	H 0 4 N 7/15	6 3 0 Z 5 B 0 5 7
G 0 6 T 1/00	3 4 0	G 0 6 T 1/00	3 4 0 A 5 C 0 6 4
7/00	1 0 0	7/00	1 0 0 C 5 L 0 9 6
7/20	3 0 0	7/20	3 0 0 B

審査請求 未請求 請求項の数26 O L (全 14 頁)

(21) 出願番号 特願2001-387569(P2001-387569)

(22) 出願日 平成13年12月20日 (2001. 12. 20)

(71) 出願人 000003049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 今井 孝志

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(72) 発明者 岩崎 一也

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

(74) 代理人 100079843

弁理士 高野 明近 (外1名)

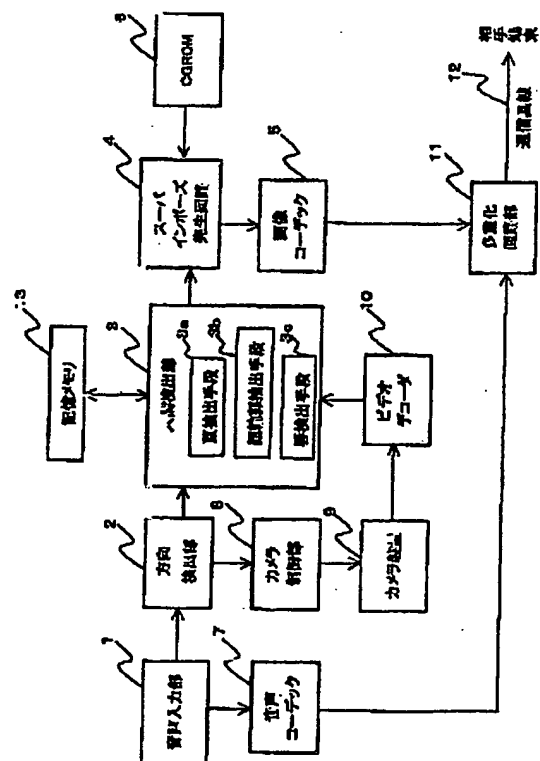
最終頁に続く

(54) 【発明の名称】 発言者識別装置及び該発言者識別装置を備えたテレビ会議システム

(57) 【要約】

【課題】 発言者を常に正確にモニタ画面に表示することが可能なテレビ会議システムを提供する。

【解決手段】 入力音声の音源方向を検出する方向検出部2によって得られた音源方向の入力画像の中から、人物を検出する人間検出部3にて人物を検出した場合に、CGR0M6とスーパーインポーズ発生回路4とにより、モニタ画面上で、検出された該人物の画像周辺に、発言者を示す所定のマーカを表示する。人間検出部3には、入力画像の中から、人物の顔の色を抽出して人物を検出する顔抽出手段3a、人物の顔の輪郭及び目乃至鼻乃至唇を検出する顔輪郭抽出手段3b、唇の動きを検出することにより発言者を検出する唇検出手段3cを備える。また、方向検出部2は、テレビ会議参加者として予め入力された音源の特徴を記憶する第1記憶メモリと新たに入力された音源の特徴を前記第1記憶メモリに記憶された音源の特徴と比較する音源特徴比較手段とを備えることもできる。



# 【特許請求の範囲】

【請求項1】 画像及び音声を入力し送受信するテレビ会議システムにおける発言者を識別するテレビ会議システムの発言者識別装置において、入力音声から音源方向を検出する方向検出手段と、入力画像内から人物を検出する人間検出手段と、表示されているモニタ画面上において前記人間検出手段にて検出された人物の画像周辺に前以って定められているマーカを表示するマーキング手段とを具備し、前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物を検出した場合に、テレビ会議における発言者として前記マーキング手段により発言者を示すマーカをマーキング表示させることを特徴とするテレビ会議システムの発言者識別装置。

【請求項2】 前記人間検出手段が、人物の顔の色を抽出して人物を検出する顔抽出手段を更に具備していることを特徴とする請求項1に記載のテレビ会議システムの発言者識別装置。

【請求項3】 前記人間検出手段が、人物の顔の輪郭及び目乃至鼻乃至口の輪郭を抽出して人物を検出する顔輪郭抽出手段を更に具備していることを特徴とする請求項1又は2に記載のテレビ会議システムの発言者識別装置。

【請求項4】 前記人間検出手段が、人物の唇が動いていることを検出する唇検出手段を更に具備していることを特徴とする請求項1乃至3に記載のテレビ会議システムの発言者識別装置。

【請求項5】 前記唇検出手段にて、人物の唇が動いていることが検出された場合に、前記マーキング手段にて発言者を示す前記マーカをマーキング表示させることを特徴とする請求項4に記載のテレビ会議システムの発言者識別装置。

【請求項6】 前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物が検出された場合であり、かつ、前記唇検出手段にて、人物の唇が動いていることが検出された場合に、前記マーキング手段にて発言者を示す前記マーカをマーキング表示させることを特徴とする請求項5に記載のテレビ会議システムの発言者識別装置。

【請求項7】 前記方向検出手段が、音源の特徴を記憶する第1記憶手段と、新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とを比較する音源特徴比較手段とを更に具備し、前記人間検出手段が、検出された人物の画像位置を記憶する第2記憶手段を更に具備し、前記音源特徴比較手段によって新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とが一致していることが検出された場合、前記第2記憶手段に記憶された前記人物の前記画像位置を読み出して、モニタ画面に表示された前記人物の画像周辺に前記マーキング手段により発言者を示す前記マーカをマーキング表示させることを特徴とする請求項1乃至6のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項8】 前記音源特徴比較手段が、前記第1記憶手段に記憶された前記音源の特徴のうち、前記新たに入力された音源の方向にある音源の特徴を読み出し、前記音源特徴比較手段によって新たに入力された音源の特徴とを比較して一致していないことが検出された場合において、前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物を検出した場合には、前記新たに入力された音源の特徴を前記第1記憶メモリに記憶し直すと共に、前記人間検出手段にて検出された該人物の画像位置を前記第2記憶メモリに記憶させることを特徴とする請求項7に記載のテレビ会議システムの発言者識別装置。

【請求項9】 前記音源特徴比較手段が、前記第1記憶手段に記憶された前記音源の特徴のうち、前記新たに入力された音源の方向にある音源の特徴を読み出し、前記音源特徴比較手段によって新たに入力された音源の特徴とを比較して一致していないことが検出された場合において、前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物を検出しなかった場合には、前記マーキング手段による発言者を示す前記マーカのマーキング表示を元の状態のままとすることを特徴とする請求項7又は8に記載のテレビ会議システムの発言者識別装置。

【請求項10】 前記第2記憶手段が、前記人間検出手段が具備している顔輪郭抽出手段により入力画像の中の人物の顔の輪郭及び／又は目乃至鼻乃至口の輪郭に関する顔特徴を抽出して記憶する顔特徴記憶手段を更に具備し、前記人間検出手段が、該顔特徴記憶手段に記憶された前記顔特徴と前記顔輪郭抽出手段により入力画像の中から新たに抽出された顔特徴とを比較する顔特徴比較手段を更に具備し、前記音源特徴比較手段によって新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とが一致していることが検出された場合、前記顔特徴比較手段において、前記顔特徴記憶手段に記憶された前記顔特徴と前記顔輪郭抽出手段によって新たに抽出された顔特徴とを比較して、一致していることが検出された場合には、前記第2記憶手段に記憶された人物の前記画像位置を読み出して、モニタ画面に表示された前記人物の画像周辺に前記マーキング手段により発言者を示す前記マーカをマーキング表示させることを特徴とする請求項7乃至9のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項11】 前記顔特徴比較手段において、前記顔特徴記憶手段に記憶された前記顔特徴と前記人間検出手段の顔輪郭抽出手段によって新たに抽出された顔特徴とを比較して、一致していないことが検出された場合、前記マーキング手段による発言者を示す前記マーカのマーキング表示を元の状態のままとすることを特徴とする請求項7乃至10のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項12】 前記顔特徴比較手段において、前記顔特徴記憶手段に記憶された前記顔特徴と前記人間検出手段の顔輪郭抽出手段によって新たに抽出された顔特徴とを比較して、一致していないことが検出された場合、前記マーキング手段による発言者を示す前記マーカのマーキング表示を元の状態のままとすることを特徴とする請求項7乃至11のいずれかに記載のテレビ会議システムの発言者識別装置。

求項10に記載のテレビ会議システムの発言者識別装置。

【請求項12】 前記方向検出手段が、音源の特徴を記憶する第1記憶手段と、新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とを比較する音源特徴比較手段とを更に具備し、前記人間検出手段が、検出された人物の画像位置を記憶する第2記憶手段と、検出された人物の唇が動いていることを検出する唇検出手段とを更に具備し、前記音源特徴比較手段によって新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とが一致していることが検出された場合、前記唇検出手段によって前記音源の特徴が一致している方向にいる人物の唇が動いていることが検出された場合には、前記第2記憶手段に記憶された人物の前記画像位置を読み出して、モニタ画面に表示された前記人物の画像周辺に前記マーキング手段により発言者を示す前記マーカをマーキング表示させることを特徴とする請求項1乃至11のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項13】 前記唇検出手段によって前記音源の特徴が一致している方向にいる人物の唇が動いていることが検出されない場合、前記マーキング手段による発言者を示す前記マーカのマーキング表示を元の状態のままとすることを特徴とする請求項12に記載のテレビ会議システムの発言者識別装置。

【請求項14】 前記方向検出手段は、入力音声が続続して入力されている継続時間を検出する時間検出手段を更に具備し、前記時間検出手段により検出された入力音声の前記継続時間が、前以って設定された所定継続時間以上に亘っていることが判定された場合において、前記人間検出手段により、入力画像の中から人物の検出を行なう動作が起動されることを特徴とする請求項1乃至13のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項15】 前記方向検出手段は、入力音声が続続して入力されている継続時間を検出する時間検出手段を更に具備し、前記時間検出手段により検出された入力音声の前記継続時間が、前以って設定された所定継続時間以上に亘っていることが判定された場合において、前記方向検出手段により、前記入力音声の音源方向の検出を行なう動作が起動されることを特徴とする請求項1乃至14のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項16】 前記方向検出手段が、入力音声の音声レベルを検出する音声レベル検出手段を更に具備し、該音声レベル検出手段により検出される入力音声の音声レベルが予め定められた所定音声レベル以上であることを検出した場合に、前記時間検出手段による前記継続時間の検出動作が開始されることを特徴とする請求項14又は15に記載のテレビ会議システムの発言者識別装置。

【請求項17】 前記人間検出手段が、入力画像内の人物を検出し、前記マーキング手段が、前記人間検出手段によって検出された人物それぞれに対して、モニタ画面に表示された該人物の画像周辺に色分け及び／又は形状分けされたマーカのマーキング表示を行なうことを特徴とする請求項1乃至16のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項18】 前記マーキング手段が、前記人間検出手段によって検出された人物それぞれに対して、モニタ画面に表示された該人物の画像周辺に色分け及び／又は形状分けをした前記マーカのマーキング表示を行なう際に、前記方向検出手段により得られた音源方向にある画像位置の人物の画像周辺には、発言者を示す色及び／又は形状の発言者マーカのマーキング表示を、また、前記方向検出手段により得られた音源方向以外にある画像位置の人物の画像周辺には、非発言者を示す色及び／又は形状の非発言者マーカのマーキング表示を、それぞれ行なうことを特徴とする請求項17に記載のテレビ会議システムの発言者識別装置。

【請求項19】 前記人間検出手段が、入力画像内の人物を検出する際に、検出された検出時刻を記憶する検出時刻記憶手段を更に具備し、該検出時刻記憶手段に記憶された前記検出時刻を、検出された前記人物の参加時刻として、モニタ画面に表示することを特徴とする請求項1乃至18のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項20】 前記人間検出手段が、入力画像内から検出された1人以上の人物の画像位置を記憶することが出来る第2記憶手段を1つ以上具備し、更に、前記人間検出手段が、前以って設定された一定周期毎に、前記入力画像内のすべての人物の検出を行ない、各人物の検出結果の画像位置をそれぞれ求める際に、前記検出結果の人物の画像位置のそれぞれと前記第2記憶手段に記憶された1人以上の人物の画像位置との比較を行なう位置比較手段を更に具備し、前記位置比較手段にて、前記第2記憶手段に記憶された人物の画像位置のいずれかが、前記検出結果の人物の画像位置のいずれにも一致していない場合には、一致していない前記第2記憶手段に記憶された画像位置の人物が退席したものととして、該人物が検出されなくなった時刻を退席時刻として記憶する退席時刻記憶手段と、逆に、前記位置比較手段にて、前記検出結果の人物の画像位置のいずれかが、前記第2記憶手段に記憶された人物の画像位置のいずれにも一致していない場合には、一致していない前記検出結果の画像位置の人物が新たに参加したものとして、該人物が新たに検出された時刻を参加時刻として記憶する参加時刻記憶手段と、を更に具備し、前記退席時刻記憶手段に記憶されている前記退席時刻を、逆に、前記参加時刻記憶手段に記憶されている前記参加時刻を、モニタ画面に表示することを特徴とする請求項1乃至19のいずれかに記載のテ

テレビ会議システムの発言者識別装置。

【請求項21】 前記方向検出手段が、入力音声の音源を検出した時刻を音源検出時刻として記憶する音源検出記憶手段を更に具備し、前記音源検出記憶手段に記憶されている前記音源検出時刻をモニタ画面に表示することを特徴とする請求項1乃至20のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項22】 少なくとも水平方向に可動可能な画像入力用のカメラ装置を更に具備し、前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けるように制御することを特徴とする請求項1乃至21のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項23】 前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けさせるように制御する際に、検出した該音源方向には前記人間検出手段による人物が検出されなかった場合、及び／又は、発言者を示す前記マークを新たにマーキング表示すべき状態が検出されなかった場合には、前記カメラ装置を動作前の位置に戻すことを特徴とする請求項22に記載のテレビ会議システムの発言者識別装置。

【請求項24】 少なくとも水平方向に可動可能な画像入力用のカメラ装置が、ズームングを行なうことができるズーム手段を更に具備し、前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けると共に前記ズーム手段によりズームングを行なうように制御することを特徴とする請求項1乃至23のいずれかに記載のテレビ会議システムの発言者識別装置。

【請求項25】 前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けさせると共に前記ズーム手段によりズームングを行なうように制御する際に、検出した該音源方向には前記人間検出手段による人物が検出されなかった場合、及び／又は、発言者を示す前記マークを新たにマーキング表示すべき状態が検出されなかった場合には、前記カメラ装置を動作前の位置に戻すことを特徴とする請求項24に記載のテレビ会議システムの発言者識別装置。

【請求項26】 請求項1乃至25のいずれかのテレビ会議の発言者識別装置を備えていることを特徴とするテレビ会議システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、テレビ会議システムの発言者識別装置及び該発言者識別装置を備えたテレビ会議システムに関する。特に、複数の参加者によりテレビ会議を行なっている際に、発言者を識別可能なマーク（目印）を付与してモニタ画面にマーキング表示する機能を有するテレビ会議システムの発言者識別装置及び該発言者識別装置を備えたテレビ会議システムに関する。

【0002】

【従来の技術】テレビ会議を行なう際に、発言者を識別し、マーク（目印）のマーキング表示を行なうテレビ会議システムの従来技術としては、例えば、特開平8-37655号公報「話者識別表示機能を有するテレビ会議システム」において開示されているように、音声の入力方向のみを識別して、発言者が存在している方向を決定し、マーキング表示を行なうものがある。即ち、特開平8-37655号公報に開示されている技術においては、複数の参加者によりテレビ会議を行なっている際に、音の発生方向を検出する音声方向検出器即ち方向検出手段からの音声方向のデータに基づいて、モニタ画面上の位置（座標）を求めて、該テレビ会議において発言している発言者（話者）の位置にカメラ装置が自動的に移動し、更に、モニタ画面に、該発言者が写し出された際には、該発言者が識別できるようなマーク（目印）を付与するように構成されているものであり、テレビ会議で発言している参加者を容易に識別することが可能であるとしている。

【0003】

【発明が解決しようとする課題】しかしながら、かかる従来の発言者識別表示機能を有するテレビ会議システムにあつては、テレビ会議の参加者の発言そのものを識別することを可能としているものではなく、何らかの音即ち物音が発生している場合に、該物音の発生方向のみに基づいて、マーク（目印）を付与してしまう構成となっている。したがって、テレビ会議の参加者の発言以外の音声や物音が発生した場合であっても、発言者識別のマーク（目印）表示がなされてしまうという問題がある。例えば、くしゃみやペンを落とした音、あるいは、マイクに物をぶつけた音など、参加者の発言以外の物音によっても、マーク（目印）のマーキング表示が行なわれてしまうため、モニタ画面に、テレビ会議にて真に必要な発言者を示す位置とは異なる位置にマーキング表示が行なわれたり、あるいは、あらぬ方向にカメラ装置の向きが移動してしまったりして、テレビ会議参加者にとっては、モニタ画面が非常に見難くなったり、紛らわしく感じられる場合が生じてしまう。

【0004】本発明は、かかる課題を解決するためになされたものであり、画像及び音声を入力し送受信するテレビ会議システムにおいて、例えば、入力音声から音源方向を検出する方向検出手段と、入力画像から人物を検出する人間検出手段とを具備し、更に、前記人間検出手段にて検出された人物の画像周辺に前以って定められている所定のマーク（目印）を表示するマーキング手段を具備し、前記方向検出手段によって得られた音源方向において、前記人間検出手段により人物が検出された場合に、始めて、テレビ会議における発言者による発声音声であると判定して、前記マーキング手段により、表示されているモニタ画面の該人物の画像周辺に発言者を示す所定のマークをマーキング表示させることを可能とす

るテレビ会議システムの発言者識別装置及び該発言者識別装置を備えたテレビ会議システムを提供せんとするものである。

【0005】

【課題を解決するための手段】本発明に係るテレビ会議システムの発言者識別装置及び該発言者識別装置を備えたテレビ会議システムは、以下の具体的な技術手段により構成されている。

【0006】第1の技術手段は、画像及び音声を入力し送受信するテレビ会議システムにおける発言者を識別するテレビ会議システムの発言者識別装置において、入力音声から音源方向を検出する方向検出手段と、入力画像内から人物を検出する人間検出手段と、表示されているモニタ画面上において前記人間検出手段にて検出された人物の画像周辺に前以って定められているマーカを表示するマーキング手段とを具備し、前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物を検出した場合に、テレビ会議における発言者として前記マーキング手段により発言者を示すマーカをマーキング表示させるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0007】第2の技術手段は、前記第1の技術手段において、前記人間検出手段が、人物の顔の色を抽出して人物を検出する顔抽出手段を更に具備しているテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0008】第3の技術手段は、前記第1又は第2の技術手段において、前記人間検出手段が、人物の顔の輪郭及び目乃至鼻乃至口の輪郭を抽出して人物を検出する顔輪郭抽出手段を更に具備しているテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0009】第4の技術手段は、前記第1乃至第3の技術手段において、前記人間検出手段が、人物の唇が動いていることを検出する唇検出手段を更に具備しているテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0010】第5の技術手段は、前記第4の技術手段において、前記唇検出手段にて、人物の唇が動いていることが検出された場合に、前記マーキング手段にて発言者を示す前記マーカをマーキング表示させるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0011】第6の技術手段は、前記第5の技術手段において、前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物が検出された場合であり、かつ、前記唇検出手段にて、人物の唇が動いていることが検出された場合に、前記マーキング手段にて発言者を示す前記マーカをマーキング表示させるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0012】第7の技術手段は、前記第1乃至第6のいずれかの技術手段において、前記方向検出手段が、音源の特徴を記憶する第1記憶手段と、新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とを比較する音源特徴比較手段とを更に具備し、前記人間検出手段が、検出された人物の画像位置を記憶する第2記憶手段を更に具備し、前記音源特徴比較手段によって新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とが一致していることが検出された場合、前記第2記憶手段に記憶された前記人物の前記画像位置を読み出して、モニタ画面に表示された前記人物の画像周辺に前記マーキング手段により発言者を示す前記マーカをマーキング表示させるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0013】第8の技術手段は、前記第7の技術手段において、前記音源特徴比較手段が、前記第1記憶手段に記憶された前記音源の特徴のうち、前記新たに入力された音源の方向にある音源の特徴を読み出し、前記音源特徴比較手段によって新たに入力された音源の特徴とを比較して一致していないことが検出された場合において、前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物を検出した場合には、前記新たに入力された音源の特徴を前記第1記憶メモリに記憶し直すと共に、前記人間検出手段にて検出された該人物の画像位置を前記第2記憶メモリに記憶させるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0014】第9の技術手段は、前記第7又は第8の技術手段において、前記音源特徴比較手段が、前記第1記憶手段に記憶された前記音源の特徴のうち、前記新たに入力された音源の方向にある音源の特徴を読み出し、前記音源特徴比較手段によって新たに入力された音源の特徴とを比較して一致していないことが検出された場合において、前記方向検出手段によって得られた音源方向に前記人間検出手段にて人物を検出しなかった場合には、前記マーキング手段による発言者を示す前記マーカのマーキング表示を元の状態のままとするテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0015】第10の技術手段は、前記第7乃至第9のいずれかの技術手段において、前記第2記憶手段が、前記人間検出手段が具備している顔輪郭抽出手段により入力画像の中の人物の顔の輪郭及び／又は目乃至鼻乃至口の輪郭に関する顔特徴を抽出して記憶する顔特徴記憶手段を更に具備し、前記人間検出手段が、該顔特徴記憶手段に記憶された前記顔特徴と前記顔輪郭抽出手段により入力画像の中から新たに抽出された顔特徴とを比較する顔特徴比較手段を更に具備し、前記音源特徴比較手段によって新たに入力された音源の特徴と前記第1記憶手段

に記憶された前記音源の特徴とが一致していることが検出された場合、前記顔特徴比較手段において、前記顔特徴記憶手段に記憶された前記顔特徴と前記顔輪郭抽出手段によって新たに抽出された顔特徴とを比較して、一致していることが検出された場合には、前記第2記憶手段に記憶された人物の前記画像位置を読み出して、モニタ画面に表示された前記人物の画像周辺に前記マーキング手段により発言者を示す前記マーカをマーキング表示させるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0016】第11の技術手段は、前記第10の技術手段において、前記顔特徴比較手段において、前記顔特徴記憶手段に記憶された前記顔特徴と前記人間検出手段の顔輪郭抽出手段によって新たに抽出された顔特徴とを比較して、一致していないことが検出された場合、前記マーキング手段による発言者を示す前記マーカのマーキング表示を元の状態のままとするテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0017】第12の技術手段は、前記第1乃至第11のいずれかの技術手段において、前記方向検出手段が、音源の特徴を記憶する第1記憶手段と、新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とを比較する音源特徴比較手段とを更に具備し、前記人間検出手段が、検出された人物の画像位置を記憶する第2記憶手段と、検出された人物の唇が動いていることを検出する唇検出手段とを更に具備し、前記音源特徴比較手段によって新たに入力された音源の特徴と前記第1記憶手段に記憶された前記音源の特徴とが一致していることが検出された場合、前記唇検出手段によって前記音源の特徴が一致している方向にいる人物の唇が動いていることが検出された場合には、前記第2記憶手段に記憶された人物の前記画像位置を読み出して、モニタ画面に表示された前記人物の画像周辺に前記マーキング手段により発言者を示す前記マーカをマーキング表示させるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0018】第13の技術手段は、前記第12の技術手段において、前記唇検出手段によって前記音源の特徴が一致している方向にいる人物の唇が動いていることが検出されない場合、前記マーキング手段による発言者を示す前記マーカのマーキング表示を元の状態のままとするテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0019】第14の技術手段は、前記第1乃至第13のいずれかの技術手段において、前記方向検出手段は、入力音声が続いて入力されている継続時間を検出する時間検出手段を更に具備し、前記時間検出手段により検出された入力音声の前記継続時間が、前以って設定された所定継続時間以上に亘っていることが判定された場合において、前記人間検出手段により、入力画像の中から

人物の検出を行なう動作が起動されるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0020】第15の技術手段は、前記第1乃至第14の技術手段のいずれかにおいて、前記方向検出手段は、入力音声が続いて入力されている継続時間を検出する時間検出手段を更に具備し、前記時間検出手段により検出された入力音声の前記継続時間が、前以って設定された所定継続時間以上に亘っていることが判定された場合において、前記方向検出手段により、前記入力音声の音源方向の検出を行なう動作が起動されるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0021】第16の技術手段は、前記第14又は第15の技術手段において、前記方向検出手段が、入力音声の音声レベルを検出する音声レベル検出手段を更に具備し、該音声レベル検出手段により検出される入力音声の音声レベルが予め定められた所定音声レベル以上であることを検出した場合に、前記時間検出手段による前記継続時間の検出動作が開始されるテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0022】第17の技術手段は、前記第1乃至第16のいずれかの技術手段において、前記人間検出手段が、入力画像内の人物を検出し、前記マーキング手段が、前記人間検出手段によって検出された人物それぞれに対して、モニタ画面に表示された該人物の画像周辺に色分け及び／又は形状分けされたマーカのマーキング表示を行なうテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0023】第18の技術手段は、前記第17の技術手段において、前記マーキング手段が、前記人間検出手段によって検出された人物それぞれに対して、モニタ画面に表示された該人物の画像周辺に色分け及び／又は形状分けをした前記マーカのマーキング表示を行なう際に、前記方向検出手段により得られた音源方向にある画像位置の人物の画像周辺には、発言者を示す色及び／又は形状の発言者マーカのマーキング表示を、また、前記方向検出手段により得られた音源方向以外にある画像位置の人物の画像周辺には、非発言者を示す色及び／又は形状の非発言者マーカのマーキング表示を、それぞれ行なうテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0024】第19の技術手段は、前記第1乃至第18のいずれかの技術手段において、前記人間検出手段が、入力画像内の人物を検出する際に、検出された検出時刻を記憶する検出時刻記憶手段を更に具備し、該検出時刻記憶手段に記憶された前記検出時刻を、検出された前記人物の参加時刻として、モニタ画面に表示するテレビ会議システムの発言者識別装置とすることを特徴とするものである。



【0025】第20の技術手段は、前記第1乃至第19のいずれかの技術手段において、前記人間検出手段が、入力画像内から検出された1人以上の人物の画像位置を記憶することができる第2記憶手段を1つ以上具備し、更に、前記人間検出手段が、前記で設定された一定周期毎に、前記入力画像内のすべての人物の検出を行ない、各人物の検出結果の画像位置をそれぞれ求める際に、前記検出結果の人物の画像位置のそれぞれと前記第2記憶手段に記憶された1人以上の人物の画像位置との比較を行なう位置比較手段を更に具備し、前記位置比較手段にて、前記第2記憶手段に記憶された人物の画像位置のいずれかが、前記検出結果の人物の画像位置のいずれにも一致していない場合には、一致していない前記第2記憶手段に記憶された画像位置の人物が退席したものとして、該人物が検出されなくなった時刻を退席時刻として記憶する退席時刻記憶手段と、逆に、前記位置比較手段にて、前記検出結果の人物の画像位置のいずれかが、前記第2記憶手段に記憶された人物の画像位置のいずれにも一致していない場合には、一致していない前記検出結果の画像位置の人物が新たに参加したものとして、該人物が新たに検出された時刻を参加時刻として記憶する参加時刻記憶手段と、を更に具備し、前記退席時刻記憶手段に記憶されている前記退席時刻を、逆に、前記参加時刻記憶手段に記憶されている前記参加時刻を、モニタ画面に表示するテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0026】第21の技術手段は、前記第1乃至第20のいずれかの技術手段において、前記方向検出手段が、入力音声の音源を検出した時刻を音源検出時刻として記憶する音源検出記憶手段を更に具備し、前記音源検出記憶手段に記憶されている前記音源検出時刻をモニタ画面に表示するテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0027】第22の技術手段は、前記第1乃至第21のいずれかの技術手段において、少なくとも水平方向に可動可能な画像入力用のカメラ装置を更に具備し、前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けるように制御するテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0028】第23の技術手段は、前記第22の技術手段において、前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けさせるように制御する際に、検出した該音源方向には前記人間検出手段による人物が検出されなかった場合、及び／又は、発言者を示す前記マーカを新たにマーキング表示すべき状態が検出されなかった場合には、前記カメラ装置を動作前の位置に戻すテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0029】第24の技術手段は、前記第1乃至第23のいずれかの技術手段において、少なくとも水平方向に

可動可能な画像入力用のカメラ装置が、ズームングを行なうことができるズーム手段を更に具備し、前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けると共に前記ズーム手段によりズームングを行なうように制御するテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0030】第25の技術手段は、前記第24の技術手段において、前記方向検出手段が検出した音源方向に、前記カメラ装置の位置を向けさせると共に前記ズーム手段によりズームングを行なうように制御する際に、検出した該音源方向には前記人間検出手段による人物が検出されなかった場合、及び／又は、発言者を示す前記マーカを新たにマーキング表示すべき状態が検出されなかった場合には、前記カメラ装置を動作前の位置に戻すテレビ会議システムの発言者識別装置とすることを特徴とするものである。

【0031】第26の技術手段は、前記第1乃至第25の技術手段のいずれかのテレビ会議システムの発言者識別装置を備えているテレビ会議システムとすることを特徴とするものである。

【0032】

【発明の実施の形態】本発明に係るテレビ会議システムの発言者識別装置及び該発言者識別装置を備えたテレビ会議システムの実施形態の一例について、以下に図面を参照しながら説明する。図1は、本発明に係る発言者識別装置を具備したテレビ会議システムの構成の一例を示すブロック構成図である。

【0033】図1に示すように、本発明に係る発言者識別装置を具備したテレビ会議システムは、入力音声から音源方向を検出する方向検出部2と、入力画像内から人物を検出する人間検出部3とを具備し、更に、モニタ画面に表示させる各種のCG (Computer Graphics) 情報を保存しているROMであるCGROM6と、該CGROM6からのCG情報を、モニタ画面に重畳させて表示させるためのスーパインポーズ発生回路4とを、識別用マーカを表示させるためのマーキング手段として具備している。図1に基づいて、本発明に係るテレビ会議システムの動作について説明する。

【0034】まず、入力された音声は、音声入力部1より電気信号からなる音声信号に変換され、方向検出部2と音声コーデック7とに送られる。方向検出部2においては、該音声信号に基づいて発言者の位置が、モニタ画面上の位置情報として検出され、カメラ制御部8および人間検出部3に対して、該位置情報が与えられる。

【0035】カメラ制御部8は、発言者の前記位置情報に基づいて、少なくとも水平方向に可動可能なカメラ装置9の旋回動作やズームング動作を行なわせることにより、音声の発生源（即ち、テレビ会議における発言者の位置）が、モニタ画面の中央に位置するように制御したり、あるいは、ズームングにより拡大表示するように制



御したりして、カメラ装置9の位置制御あるいはズーム制御を行なう。

【0036】ここで、カメラ制御部8は、方向検出部2により検出された音声の発声方向（音源方向）にカメラ装置9の位置を向けるように旋回制御させたり、ズーム制御をさせたりする際に、方向検出部2により検出された音源方向に人間検出部3が人物を検出できなかった場合、及び／又は、例えば、テレビ会議の参加者の音声（音源）の特徴と一致していない特徴の発声音声である場合などのごとく、発言者を特定することができなかった場合（即ち、発言者を示すマーカを新たにマーキング表示すべき状態を検出できなかった場合）には、動作前の元のカメラ装置9の位置に戻すことも可能である。

【0037】あるいは、前述のごとく、カメラ装置9はズーム手段を具備しており、カメラ制御部8は、ユーザの指示により、又は、方向検出部2により検出された音源方向に、カメラ装置9の位置を向けると共に、自動的にカメラ装置9のズームングを行なわせるように制御することも可能であり、また、方向検出部2により検出された音源方向に人間検出部3が人物を検出できなかった場合、及び／又は、発言者を特定することができなかった場合（即ち、発言者を示すマーカを新たにマーキング表示すべき状態を検出できなかった場合）には、動作前の元のカメラ装置9の位置に戻すと共にカメラ装置9のズームング状態を動作前の状態に戻すことも可能である。

【0038】撮像装置としてのカメラ装置9においては、撮像された映像情報を、映像信号として電気信号に変換する。変換された該映像信号はビデオデコード10によってデジタル処理され、人間検出部3に映像データとして送られる。人間検出部3においては、方向検出部2から入力される前記位置情報とビデオデコード10から入力される前記映像データとに基づいて、前記映像データにおける前記位置情報が示す位置に、人物の撮像画像データが存在しているか否かの検出を行なう。

【0039】前記位置情報が示す位置に、人物の撮像画像データが存在していることが検出された場合にあっては、マーキング手段を提供するスーパーインポーズ発生回路4において、各種のCG（Computer Graphics）情報を保存しているROMであるCGROM6から、発言者を示すマーカ（目印）として予め定められているマーカデータを読み出し、前記映像データ上における前記位置情報と発言者である人物の撮像画像データとから演算されたマーカ表示位置（即ち、発言者の人物の画像周辺位置）に、読み出された発言者を示す前記マーカデータを前記映像データ上に重ね合わされた合成映像データが作成される。

【0040】作成された合成映像データは、画像コーデック5によって圧縮符号化処理された符号化画像データとされ、一方、音声入力部1からの音声信号が、音声コ

ーデック7にて圧縮符号化された符号化音声データとされ、該符号化画像データと該符号化音声データとが、多重化回路部11にて多重化されて、通信回線12を通して、相手端末に送られる。

【0041】ここで、マーキング手段であるスーパーインポーズ発生回路4にて作成された前記合成映像データ（即ち、カメラ装置9で撮像され、ビデオデコード10でデジタル処理が施された映像データと、発言者を示すマーカデータとが合成された合成映像データ）が更に画像コーデック5により圧縮符号化された状態の前記符号化画像データが、受信側の相手端末における画像コーデックにより復号化されて、テレビ会議システムにおけるモニタ画面に画面表示されている画像の一例の概念図を、図2に示す。図2において、101は、テレビ会議の参加者のうち、発言者を示し、102は、テレビ会議の参加者で現在発言していない人物を示し、100は、発言者であることを示すマーカ（図2においては、矢印形状のマーク）である。

【0042】人間検出部3は、カメラ装置9からの映像情報がビデオデコード10によりデジタル処理が施された映像データの中から色情報を用いて人間の顔の色を抽出することにより、顔領域を特定（抽出）して、映像データの中に含まれている人物を検出することができる顔抽出手段3aを備えている。また、該人間検出部3は、カメラ装置9からの映像情報がビデオデコード10によりデジタル処理が施された映像データの中から、人物の顔の輪郭や目乃至鼻乃至口といった個々の特徴を有する輪郭を抽出することによって、映像データの中に含まれている人物を検出することができる顔輪郭抽出手段3bも備えている。

【0043】更には、後述するように、該人間検出部3は、カメラ装置9からの映像情報がビデオデコード10によりデジタル処理が施された映像データの中から、人物の唇が動いているか否かを検出することができる唇検出手段3cも備えている。かかる唇検出手段3cにより、参加者の人物の唇が動いていることが検出された場合、方向検出部2の結果如何によらず、あるいは、方向検出部2によって得られた音源方向に人間検出部3にて人物を検出した場合であって、かつ、前記唇検出手段3cにて該人物の唇が動いていることが検出された場合には、該人物が発言していると見なして、モニタ画面上の発言者である該人物の画像周辺に、発言者を示すマーカをマーキング表示させることとしても良いし、更に、方向検出部2により検出された音源方向に一致する方向にいる人物の唇が動いていることが、唇検出手段3cにより検出された場合であって、更に、方向検出部2にて発声音声（音源）の特徴とテレビ会議参加者の発声音声（音源）の特徴とが一致する場合において、発言者を示すマーカをモニタ画面にマーキング表示させることとしても構わない。

【0044】また、図1に示す方向検出部2と人間検出部3とは、図3のような構成にすることも可能である。ここに、図3は、本発明に係る発言者識別装置を具備したテレビ会議システムを構成する方向検出部と人間検出部とその他の構成例を示すブロック構成図である。即ち、図3に示すごとく、方向検出制御部2'は、図1に示す方向検出部2以外に、更に、入力音声（音源）の特徴を抽出する音源特徴抽出部21と、抽出された入力音声（音源）の特徴を特徴データとして記憶する第一の記憶手段である第1記憶メモリ22と、方向検出部2を介して入力されて、音源特徴抽出部21により抽出された音源の特徴と第1記憶メモリ22に記憶されているすべての前記特徴データとを比較照合する音源特徴比較部23とを備えている。

【0045】一方、人間検出制御部3'は、図1に示す人間検出部3以外に、更に、人間検出部3において、ビデオデコード10からの映像データの中に人物が検出された場合の人物の画像位置を記憶するための第二の記憶手段を提供する第2記憶メモリ32と、方向検出部2が示す音声の発声方向（音源方向）の位置に相当する第2記憶メモリ32における映像データの画像位置を算出する演算を行なう位置演算部31とを備えている。

【0046】図3に示す方向検出制御部2'と人間検出制御部3'とにおいては、まず、テレビ会議が始まるに先立って、第1記憶メモリ22と第2記憶メモリ32とに、それぞれ、テレビ会議の各参加者の発声音声（音源）の特徴を示す特徴データと各参加者の画像位置とを予め登録する。ここで、発言者の発声音声が、音声入力部1より電気信号の音声信号に変換されて、方向検出制御部2'に入力され、図1と同一の機能を果たす方向検出部2を介して、音源特徴抽出部21に送られる。音源特徴抽出部21においては、発言者の音声信号（音源）の特徴を抽出し、第1記憶メモリ22に特徴データとして格納する。

【0047】一方、図1と同一の人物検出機能を果たす人間検出部3においては、方向検出制御部2'から入力される発声音声方向（音源方向）の位置に相当するビデオデコード10の映像データ内の画像位置に人物の映像を検出した場合には、該人物の画像位置を第2記憶メモリ32に格納する。

【0048】第1記憶メモリ22と第2記憶メモリ32とに、テレビ会議参加者に関する音声（音源）の特徴と人物の画像位置とを予め設定した後、テレビ会議が始まると、音源特徴比較部23においては、音源特徴抽出部21から送られてくる新たな発声音声（音源）の特徴と、第1記憶メモリ22に記憶されているすべての音声の特徴データとを比較照合し、記憶されているすべての音声の前記特徴データの中に、新たな前記発声音声（音源）の特徴と一致する音声が存在していることが検出された場合には、本テレビ会議の参加者の発言と判断し、

位置演算部31を介して該人物の画像位置を第2記憶メモリ32から読み出して、モニタ画面上における該発言者の人物の画像周辺に、発言者であることを示すマークを、スーパーインポーズ発生回路4を介して画面表示する。

【0049】かくのごとく、発声音声である音源の特徴に基づいて、テレビ会議の参加者の発言と判断される場合にあっては、図1に示す人間検出部3と同一の機能を果たす図3における人間検出部3は、何ら動作をする必要はなく、人間検出部3は起動されることがないものとすることができる。

【0050】一方、もし、音源特徴抽出部21から送られてくる前記発声音声（音源）の特徴が、第1記憶メモリ22に記憶されているすべての音声の特徴データと一致していないことが音源特徴比較部23において判明した場合には、図1に示す人間検出部3と同一の機能を果たす図3における人間検出部3が起動されて、該人間検出部3において、ビデオデコード10からの映像データの中から人物検出を行なう。即ち、該人間検出部3において、位置演算部31を介して得られた前記発声音声方向（音源方向）の位置に、人物の映像が存在しているか否かが判別されることにより、該発声音声（音源）が、本テレビ会議の参加者の発言であるか否かを判定する。

【0051】人間検出部3において、前記発声音声方向（音源方向）の位置に人物が存在していないと判定された場合には、該発声音声（音源）が、本テレビ会議には無関係の雑音と見なして、何ら処理を行なうことなく、元の状態のままとし、一方、該発声音声方向（音源方向）の位置に人物が存在していると判定された場合には、発言者が位置を移動して発言しているものと見なして、位置演算部31を介して算出されている参加者の画像位置情報を、新たに第2記憶メモリ32に登録すると共に、第1記憶メモリ22にも、該発声音声（音源）の特徴データを再登録する。

【0052】従来の技術においては、新たな音声が発生された際には、発言者を示すマーク（目印）をモニタ画面にマーキング表示させたり、カメラ装置を旋回させて、該音声の発生元である発言者がモニタ画面内に収まるように撮像せんとしている。しかしながら、発言者以外の何らかの物音が発生した場合においても、全く同様に、発言者を示すマーク（目印）がモニタ画面にマーキング表示されたり、あるいは、カメラ装置が物音の発生方向に旋回されてしまっていた。

【0053】本発明に係る発言者識別手段を備えたテレビ会議システムにおいては、人間検出部3にて人物が検出されなかった場合には、発言者を示すマークを新たにマーキング表示させることも行なわれないし、更に、図1に示すカメラ制御部8を制御して、カメラ装置9の位置やズーム状態が、元の状態に戻るよう復帰指令を送出している。即ち、たとえ、カメラ装置9の位置が

一旦音源方向に旋回されたとしても、図示はしていないが、カメラ制御部8が旋回制御される前の位置情報やズーム情報、例えば、図3に示す第2記憶メモリ32に保存されていることにより、カメラ装置9の位置やズーム状態を元の状態に戻す復帰指令の送出が可能とされている。而して、たとえ、参加者の発言以外の物音に反応して、カメラ装置9が旋回してしまった場合であっても、発言者を示すマーカ（目印）の位置、あるいは、カメラ装置9の位置を、元の位置に戻すことができる。

【0054】また、第2記憶メモリ32が、人間検出部3が具備している顔輪郭抽出手段3bにより人物の顔の輪郭及び目乃至鼻乃至口の輪郭に関する個々の人物の顔特徴を抽出して記憶する顔特徴記憶手段を更に具備している場合においては、人間検出部3が、該顔特徴記憶手段に記憶された前記顔特徴と顔輪郭抽出手段3bにより映像データの中から新たに抽出された顔特徴とを比較する顔特徴比較手段を更に具備し、音源特徴比較部23によって新たに入力された音源の特徴と第1記憶メモリ22に記憶された前記特徴データ（音源の特徴）とが一致していることが検出された場合、前記顔特徴比較手段において、前記顔特徴記憶手段に記憶された前記顔特徴と人間検出部3の顔輪郭抽出手段3bによって新たに抽出された顔特徴とを比較して、一致していることが検出された場合には、第2記憶メモリ32に記憶された人物の画像位置を読み出して、該画像位置の前記人物の画像周辺に発言者を示すマーカをマーキング表示させたり、カメラ装置9を前記人物の位置に旋回させたり、ズームして拡大表示させることも可能である。

【0055】ここで、前記顔特徴比較手段において、前記顔特徴記憶手段に記憶された前記顔特徴と顔輪郭抽出手段3bによって新たに抽出された顔特徴とを比較して、一致していないことが検出された場合、発言者を示す前記マーカのマーキング表示やカメラ装置9の位置やズーム状態を元の状態のままとすることとする。

【0056】更に、図1に示す方向検出部2は、図4のような構成にすることも可能である。ここに、図4は、本発明に係る発言者識別装置を具備したテレビ会議システムを構成する方向検出部の更なる他の構成例を示すブロック構成図である。即ち、図4に示すように、方向検出制御部2'は、図1に示す方向検出部2以外に、更に、所定の音声レベルを上回る発声音声（音源）を検出する音声レベル検出部24と、かかる所定の音声レベルを上回る発声音声が予め設定されている所定継続時間以上に亘って継続していることを検出する時間検出手段を提供するタイマ部25とを備えている。

【0057】図4において、入力された発声音声は、音声入力部1において、音声信号に変換されて、方向検出制御部2'に入力され、音声レベル検出部24において発声音声の音声レベルが測定される。参加者からの発言

として、該発声音声が前記所定の音声レベル以上の音声信号であることが検出された場合には、タイマ部25のタイマが起動されて、経過時間の計数が開始され、予め設定されている前記所定継続時間が経過した場合には、所定の継続時間以上に亘って、入力音声（音源）が継続している状態にあり、タイマ部25からの出力信号を、方向検出部2に対して送出する。

【0058】ここで、予め設定されている前記所定継続時間が経過する前に、参加者からの発言が終了して、音声レベル検出部24において、前記所定の音声レベル以上の発声音声が検出されなくなると、タイマ部25はタイマの計数を停止され、タイマ部25からの出力信号は発生しなくなる。図1と同様の機能を果たす方向検出部2は、タイマ部25からの出力信号が入力されている場合にあって、始めて、入力された発声音声（音源）が、本テレビ会議の参加者の発言に基づく音声信号であるか否かを判断する動作が起動されることになる。

【0059】例えば、くしゃみのような短い時間の音声情報は発言ではないので、図1に示すカメラ装置9を、短い時間の該音声情報の位置に旋回させたり、あるいは、モニタ画面上に該音声情報の位置を発言者を示すようにマーキング表示させることは無駄である。本発明における実施例に示すように、かかる短い時間の音声情報の場合においては、タイマ部25からの出力信号は出力されることはなく、而して、方向検出部2としては、発声音声方向（音源方向）の位置を示す位置情報を検出する動作が起動されずに、カメラ装置9の旋回などの制御動作や、あるいは、人間検出部3における人物検出動作も起動されない状態に設定されている。

【0060】また、図4においては、かかる短い時間の音声情報の場合に、方向検出部2が起動されないことにより、人間検出部3も起動されない旨を説明しているが、かかる場合に限らず、直接、タイマ部25から人間検出部3へも出力信号が供給されていて、かかる短い時間の音声情報の場合には、方向検出部2を介することなく、直接、人間検出部3を起動させない状態とすることも可能である。

【0061】なお、以上に説明のごとく、図4においては、タイマ部25を起動させる条件として、予め定められている所定の音声レベル以上のレベルにある発声音声がある場合を条件としているが、かかる所定の音声レベル以上のレベルにある発声音声であるか否かの如何に関わらず、識別可能な何らかの音声が続いて、所定継続時間以上に亘っていることが検出された場合であっても構わない。

【0062】また、前述のごとく、図3に示す人間検出部3は、方向検出部2からの発声音声方向（音源方向）の位置が示す映像データ上の位置に存在する人物の有無を検出するだけではなく、テレビ会議に先立って、予め、テレビ会議への参加者全員の画像位置の検出を行な

い、図3に示す第2記憶メモリ32に登録しておくようにすることが可能とされている。而して、例えば、図5に示すように、参加者のうち、発言者と非発言者とのマーカの表示を色分けしたり、あるいは、形状分けしたりして、変化させて、モニタ画面上にマーキング表示を行なわせることも可能である。

【0063】ここに、図5は、テレビ会議参加者のうち、発言者用と非発言者用のマーカのマーキング表示を行なう場合のモニタ画面表示の一例を示す概念図である。図5においては、発言者105が発言していることを示す発言者マーカ103は、非発言者106を示す非発言者マーカ104とは、例えば、異なる色を用いてマーキング表示している一例を示している。

【0064】即ち、テレビ会議参加者のうち、発言者105が発言を行なった場合、方向検出部2にて、発言者105の音源方向を示す位置情報が検出され、予め記憶メモリ32に記憶されている画像位置と比較して、該音源方向に人物の存在を確認することによって、発言者105を識別し、発言していない非発言者106を示す非発言者マーカ104とは異なる色の発言者マーカ103によって表示している。

【0065】而して、たとえ、テレビ会議の参加者全員をモニタ画面に画面表示している状態であっても、モニタ画面の参加者各自毎に即ち人物の画像周辺位置に、参加者全員に対してそれぞれ色違いの発言者マーカ103と非発言者マーカ104とのマーキング表示を行ない、発言者か否かを容易に識別可能とすることができる。ここに、マーキング表示する前記マーカとしては、発言者と非発言者との識別用のみに限るものではない。例えば、テレビ会議の司会者と一般参加者とオブザーバとを容易に識別可能とするように、色分け及び／又は形状分け及び／又は模様分けすることとしても良く、テレビ会議の実施に有用な如何なる識別情報でも、モニタ画面に重畳表示されるマーカにより提供することが可能である。

【0066】また、人間検出部3を構成する人間検出制御部3'は、音声が入力された時だけ、人物検出動作を行なうのではなく、常時、定期的に、人間検出部3にて人物の検出動作を行なうこともできる。而して、ビデオデコーダ10からの映像データの中の各画像位置で、人物を最初に検出した時刻を、図示していない時計回路から読み出し、検出された該人物の検出時刻を、テレビ会議への参加時刻として、図3に示す第2記憶メモリ32の検出時刻記憶部に記憶させると共に、スーパインポーズ発生回路4により、該検出時刻記憶部から読み出した前記検出時刻を重ね合わせて画面表示させることにより、モニタ画面には、人物の検出結果として、該人物がテレビ会議に参加した参加時刻を表示することができる。

【0067】あるいは、人間検出部3を構成する人間検

出制御部3'は、ビデオデコーダ10からの映像データ即ち入力画像内から検出される人物の画像位置を1人以上記憶することができる1つ以上の第2記憶メモリ32を具備し、更に、予め設定されている一定周期毎に、定期的に、前記入力画像内のすべての人物の検出を行ない、各人物の検出結果の画像位置をそれぞれ求める際に、前記検出結果の人物の画像位置のそれぞれと、第2記憶メモリ32に記憶された1人以上の人物の画像位置とを比較する位置検出手段を更に具備しており、該位置検出手段にて、前記検出結果の人物の画像位置のいずれかが、第2記憶メモリ32に記憶された人物の画像位置のいずれにも一致していない場合には、前記検出結果の画像位置の人物が新たにテレビ会議に参加したものを見なし、人間検出部3により前記検出結果の画像位置の人物が新たに検出された時刻を参加時刻として、図3に示す第2記憶メモリ32の参加時刻記憶部に記憶させると共に、該参加時刻記憶部に記憶された前記参加時刻を、モニタ画面に表示させることとしても良い。

【0068】例えば、図6は、あるテレビ会議の参加状況を示すモニタ画面表示の一例を示す概念図である。図6(A)において、109及び110は、9:00にテレビ会議に参加した参加者を示すものであり、それぞれの人物が検出された検出時刻即ち参加時刻を、モニタ画面上の映像データに重ね合わせて、該参加者の人物画像位置の画像周辺に、参加時刻表示107及び108が表示されている。

【0069】例えば、図6(B)は、図6(A)から、しばらく時間が経過したテレビ会議の参加状況を示しており、モニタ画面に参加時刻表示111が新たに追加されて表示されているように、参加者112が9:30にテレビ会議に参加したことがわかる。かかる参加時刻の表示は、図3に示す第2記憶メモリ32の前記検出時刻記憶部又は前記参加時刻記憶部に記憶させることにより、モニタ画面には、テレビ会議が終了するまで、該人物の参加時刻を引き続き表示させることができる。

【0070】また、人間検出制御部3'は、ビデオデコーダ10からの映像データ即ち入力画像内から検出される人物の画像位置を1人以上記憶することができる1つ以上の第2記憶メモリ32を具備し、更に、予め設定されている一定周期毎に、定期的に、前記入力画像内のすべての人物の検出を行ない、各人物の検出結果の画像位置をそれぞれ求める際に、前記検出結果の人物の画像位置のそれぞれと、第2記憶メモリ32に記憶された1人以上の人物の画像位置とを比較する位置検出手段を更に具備しており、該位置検出手段にて、第2記憶メモリ32に記憶された人物の画像位置のいずれかが、前記検出結果の画像位置のいずれにも一致していない場合には、第2記憶メモリ32に記憶された一致していない画像位置の人物が退席したものを見なし、人間検出部3により該人物が検出されなくなった時刻を退席時刻として、図

3に示す第2記憶メモリ32の退席時刻記憶部に記憶させると共に、該退席時刻記憶部に記憶された前記退席時刻を、モニタ画面に表示させることができる。

【0071】即ち、今まで人物検出がされていた画像位置即ち画像領域において人間検出部3による人物の検出がされなくなった場合、図示していない時計回路より人物が検出されなくなった時刻を読み出し、スーパインポーズ発生回路4により重ね合わせ表示させてモニタ画面に表示させていることになる。而して、モニタ画面には、人物が検出されなくなった時刻を、該参加者が退席した退席時刻として、該参加者を示していた人物の画像位置の画像周辺に画面表示することができる。また、かかる人物が検出されなくなった退席時刻は、図3に示す第2記憶メモリ32の前記退席時刻記憶部に記憶させることにより、モニタ画面には、テレビ会議が終了するまで、該人物の退席時刻を引き続き表示させることができる。

【0072】更に、人間検出部3は、発言者の発声音声方向（音源方向）による位置情報とカメラ装置9からの映像情報に基づく映像データとにより、音声を発声した発言者の人物を検出した場合、図示していない時計回路より時刻を読み出し、スーパインポーズ発生回路4より読み出された時刻を重ね合わせて表示することもできる。即ち、方向検出制御部2'が、入力音声の音源を検出した時刻を音源検出時刻として、図3に示す第1記憶メモリ22の音源検出記憶部に記憶すると共に、モニタ画面には、該音源検出記憶部に記憶されている前記音源検出時刻を、各参加者の人物画像位置の画像周辺に表示させることができる。而して、参加者各々が発言した発言時刻を、モニタ画面に表示することができる。

【0073】また、人間検出部3は、人物の顔の色や顔の輪郭などによる人物検出機能と共に、前述のごとく、人物の唇が動いていることを検出する唇検出手段3cを更に付与させることも可能としている。

【0074】かかる場合においては、前述のごとく、発声音声が音声入力部1に入力された場合、まず、人間検出部3の人物検出機能により、発声音声方向（音源方向）の位置にいる人物が検出される。その後、唇検出手段3cにより、検出された人物の唇が動いているか否かが検出される。唇が動いていることが検出できれば、入力された発声音声方向（音源方向）にいる当該人物の発言による発声音声であると判断し、カメラ装置9の旋回動作やズーム動作、更には、および発言者を示すマーカのマーキング表示を行なう。逆に、唇検出手段3cにより、検出された人物の唇が動いていないことが検出された場合は、入力された発声音声は、入力された発声音声方向（音源方向）にいる当該人物の発言による発声音声ではないと判断し、カメラ装置9の旋回動作やズーム動作及び発言者を示すマーカのマーキング表示を行なわない。

【0075】更には、図3に示すごとく、方向検出制御部2'が、音源特徴抽出部21により抽出された音源の特徴を記憶する第1記憶メモリ22と、新たに入力された音源の特徴と第1記憶メモリ22に記憶された前記音源の特徴とを比較する音源特徴比較部23とを具備して、人間検出制御部3'が、検出された人物の画像位置を記憶する第2記憶メモリ32を具備している場合において、人間検出制御部3'が、更に、検出された人物の唇が動いていることを検出する唇検出手段3cを具備している場合にあっては、音源特徴比較部23によって新たに入力された音源の特徴と第1記憶メモリ22に記憶された前記音源の特徴とが一致していることが検出された場合で、かつ、前記音源の特徴の一致が検出された方向にいる人物の唇が動いていることが、唇検出手段3cによって検出された場合に、始めて、第2記憶メモリ32に記憶された人物の前記画像位置を読み出して、該画像位置の前記人物の画像周辺に発言者を示す前記マーカをマーキング表示させたり、カメラ装置9の位置を発言者の方向に向けさせたり、ズームさせることも可能である。

【0076】かかる場合において、前記音源の特徴が一致している方向にいる人物の唇が動いていることが、唇検出手段3cによって検出されない場合にあっては、ごく短時間の発言であったものと見なして、発言者を示す前記マーカをマーキング表示させる位置やカメラ装置9の位置を、元の状態のままとすることも可能である。

【0077】

【発明の効果】以上に説明したごとく、本発明に係るテレビ会議システムの発言者識別装置及び該発言者識別装置を備えたテレビ会議システムによれば、例えば、くしゃみや物を落とした音、マイクに物をぶつけた音など参加者の発言以外の音の発生によって、誤って、該発言以外の音の発生方向にカメラ装置が旋回してしまったり、発言者を示すマーカをマーキング表示させたりすることを防止することが可能であり、発言している参加者を常に正確にモニタ画面に表示することが可能となり、而して、快適にテレビ会議を行なうことが可能なテレビ会議システムを提供することができる。

【0078】更には、たとえ、テレビ会議の参加者全員をモニタ画面に画面表示した場合であっても、発言者や非発言者を識別可能なマーカを重畳させて画面表示させたり、あるいは、テレビ会議の司会者や一般の参加者あるいはオブザーバなどの識別が視覚的に容易なマーカを重畳させて画面表示させることも可能であり、而して、スムーズにテレビ会議を行うことが可能なテレビ会議システムを提供することができる。

【0079】更には、テレビ会議の参加者が会議に参加した参加時刻、退席した退席時刻、発言した発言時刻などのテレビ会議の会議経過時刻を表示することが可能なテレビ会議システムを提供することができる。

## 【図面の簡単な説明】

【図1】本発明に係る発言者識別装置を具備したテレビ会議システムの構成の一例を示すブロック構成図である。

【図2】本発明に係る発言者識別装置を具備したテレビ会議システムにおいて、モニタ画面に画面表示された画像の一例を示す概念図である。

【図3】本発明に係る発言者識別装置を具備したテレビ会議システムを構成する方向検出部と人間検出部とその他の構成例を示すブロック構成図である。

【図4】本発明に係る発言者識別装置を具備したテレビ会議システムを構成する方向検出部の更なる他の構成例を示すブロック構成図である。

【図5】テレビ会議参加者のうち、発言者用と非発言者用のマーカのマーキング表示を行なう場合のモニタ画面表示の一例を示す概念図である。

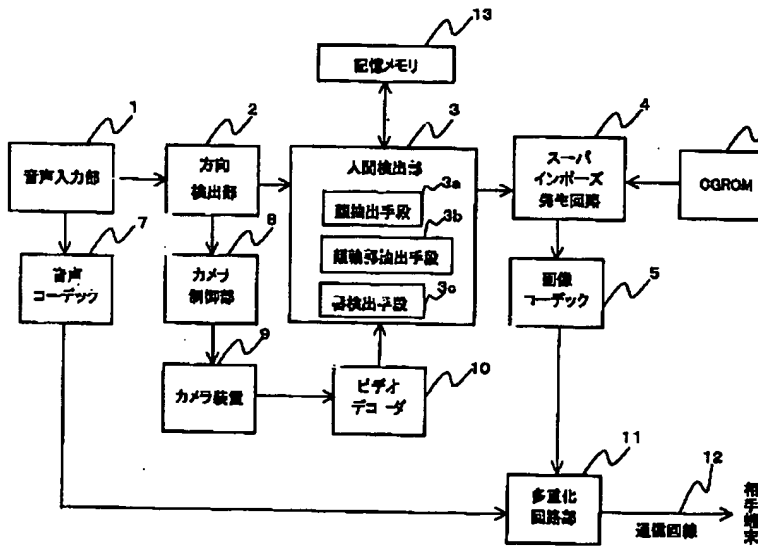
【図6】あるテレビ会議の参加状況を示すモニタ画面表

示の一例を示す概念図である。

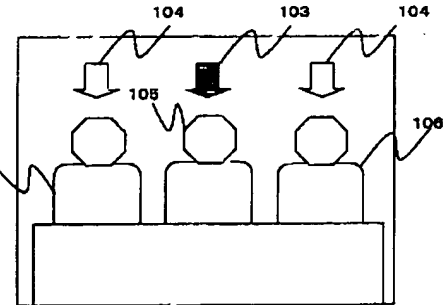
## 【符号の説明】

1…音声入力部、2…方向検出部、2'、2''…方向検出制御部、3…人間検出部、3'…人間検出制御部、3a…顔抽出手段、3b…顔輪郭抽出手段、3c…唇検出手段、4…スーパインポーズ発生回路、5…画像コーデック、6…CGROM、7…音声コーデック、8…カメラ制御部、9…カメラ装置、10…ビデオデコーダ、11…多重化回路部、12…通信回線、13…記憶メモリ、21…音源特徴抽出部、22…第1記憶メモリ、23…音源特徴比較部、24…音声レベル検出部、25…タイマ部、31…位置演算部、32…第2記憶メモリ、100…マーカ、101…発言者、102…参加者、103…発言者マーカ、104…非発言者マーカ、105…発言者、106…非発言者、107、108…参加時刻表示、109、110…参加者、111…参加時刻表示、112…参加者。

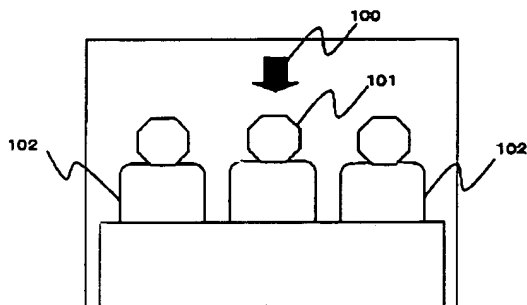
【図1】



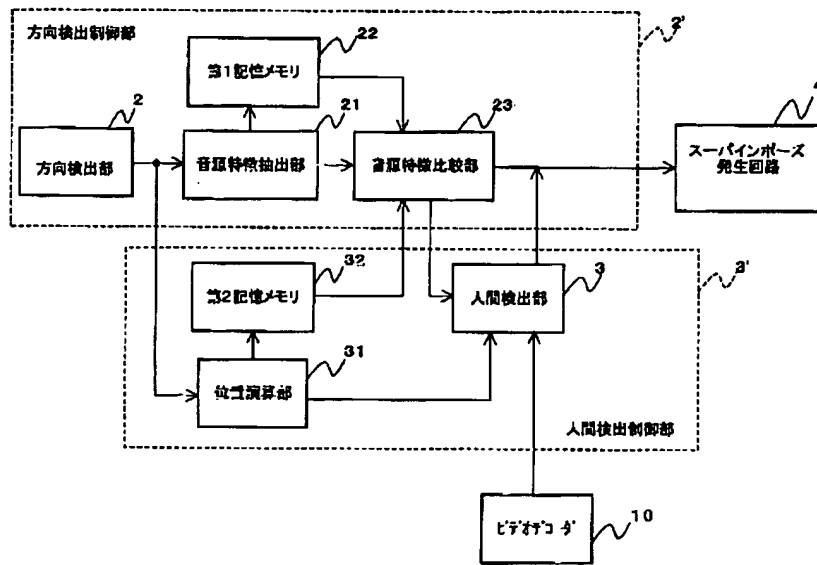
【図5】



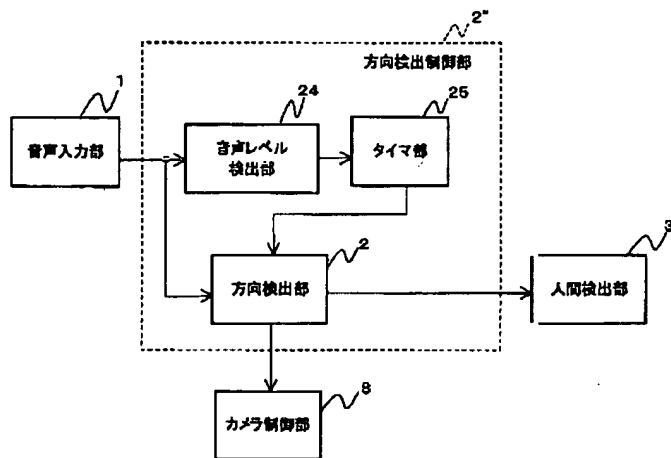
【図2】



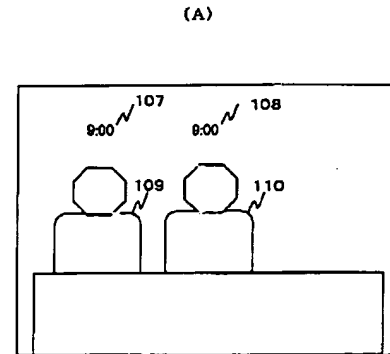
【図3】



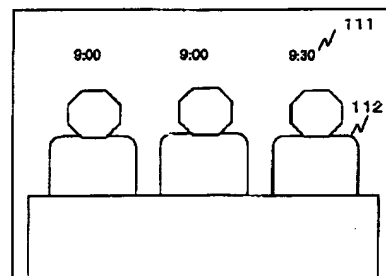
【図4】



【図6】



(B)



フロントページの続き

Fターム(参考) 5B057 AA20 BA02 BA17 CA01 CA08  
 CA16 CB01 CB08 CB12 CB16  
 CE08 CH18 DA07 DA08 DA16  
 DB02 DB06 DB09 DC16 DC25  
 DC36  
 5C064 AA02 AC04 AC06 AC08 AC09  
 AC12 AC13 AC15 AC16 AC18  
 AC22 AD02  
 5L096 AA02 AA06 BA20 CA02 DA02  
 FA06 FA15 GA38 HA03